Nonparametric Inference for a Triangular System of Equations for Quantile Regression

Yubin Kim and Sungwon Lee

In this study, we consider nonparametric estimation and inference for quantile regression (QR) with endogenous regressors. We extend the semiparametric triangular model for QR in Lee (2007) to a nonparametric one, and the identification of the structural parameters is achieved via a control function approach. Based on the identification result, we propose the use of the penalized sieve minimum distance procedure of Chen and Pouzo (2015) and develop an asymptotic theory. The inferential theory is valid regardless of whether or not the functional of the structural parameter is $\sqrt{n}$ -estimable, where n denotes the number of observations. We also establish the asymptotic theory for sieve quasi-likelihood ratio test statistics, enabling us to avoid estimating the asymptotic variance. A Monte Carlo simulation study shows that the proposed estimator performs well in finite samples.

Keywords: Quantile regression, Endogeneity, Nonparametric simultaneous equations model, Sieve estimation, Sieve quasi-likelihood ratio test statistics.
JEL Classification: C13, C14, C31.

Yubin Kim, Brown University and Sogang University; (E-mail): yubinkim@u.sogang.ac.kr

Sungwon Lee (Corresponding author), Department of Economics, Sogang University, 35 Baekbeom-ro, Mapo-gu, Seoul, South Korea 04107; (E-mail): sungwonglee@sogang.ac.kr

[**Seoul Journal of Economics** 2025, Vol. 38, No.1]
DOI: 10.22904/sje.2025.38.1.001

I. Introduction

Since the publication of the seminal paper by Koenker and Bassett (1978), quantile regression (QR) has become a popular approach in empirical studies in the field of economics. Using QR, one can investigate the heterogeneous effects of covariates on outcomes across their distribution. Such heterogeneous effects, if they exist, can provide a rich set of policy implications that can be difficult to provide with the mean regression (*e.g.*, Bitler *et al.* 2006, 2008). However, given that endogeneity is prevalent in observational settings and causes inconsistency of QR estimators, many studies have considered instrumental variable estimation within the QR framework.[1]

The main goal of this paper is to provide a tractable nonparametric estimation and inference approach for QR with endogenous regressors. Specifically, we develop inference methods for the nonparametric QR model of Lee (2022) , who extended the semiparametric QR model with the endogenous regressors of Lee (2007). In this work , we use a triangular system of equations, in which the reduced-form equations for endogenous regressors are specified. Based on the triangular system, we apply a control function approach to identifying the structural functions, as in Newey *et al.* (1999), Lee (2007), Chernozhukov *et al.* (2015), and Lee (2022).

Our estimation strategy relies on the methods of sieves, which allow for flexible modeling and provide a tractable estimation strategy (*cf.* Chen 2007). Specifically, we consider the penalized sieve minimum distance (PSMD) estimator of the unknown infinite-dimensional parameter proposed by Chen and Pouzo (2015) and develop the asymptotic theory, including consistency, convergence rates, and asymptotic normality.

The multiple-step estimation procedure is commonly used when the model parameters are identified via a control function approach (*e.g.*, Newey *et al.* 1999; Das *et al.* 2003; Newey (2009); Chernozhukov *et al.* 2015). In such cases, the issue of generated regressors must be considered, and there are several important papers in the literature that address this issue (*e.g.*, Ackerberg *et al.* 2012; Mammen *et al.*

[1] Please refer to Chernozhukov and Hansen (2013) for an excellent review on QR with endogeneity.

2012; Hahn and Ridder 2013; Hahn *et al.* 2018). In particular, the two-step sieve estimation approach developed by Hahn *et al.* (2018) can be adapted to our context. However, there are several advantages of our approach compared to that of Hahn *et al.* (2018). First, our inference theory is valid regardless of whether or not the functional of interest is $\sqrt{n}$-estimable, where $\sqrt{n}$ denotes the number of observations. A functional that is $\sqrt{n}$-estimable is called a "regular functional." If it is not $\sqrt{n}$-estimable, we call it an "irregular functional." While it is already challenging to verify whether a functional of interest is regular or irregular in actual practice, this may even be worsened when the model is highly nonlinear. The general inference theory developed in Chen and Pouzo (2015) can be applied to both regular and irregular functionals. Consequently, the inference methods in this paper are also applicable to both regular and irregular functionals and have a wide applicability. Meanwhile, Hahn *et al.* (2018) focused on inference on regular functionals based on a two-step estimation procedure.

Second, we establish the asymptotic distribution of sieve quasi-likelihood ratio (QLR) test statistics, which do not require the estimation of the asymptotic variance of the specific functional's estimator. It is well known that the asymptotic variance of a QR estimator typically involves the conditional density function of the outcome variable given covariates. Although this infinite-dimensional parameter may be cumbersome to estimate in practice, one can circumvent the estimation of the asymptotic variance by using some proper QLR statistics.

Finally, when we use a semiparametric model, the sieve estimators of finite-dimensional parameters become semiparametrically efficient. Although our main focus is on nonparametric QR models, one can impose some semiparametric structure on the model to avoid the curse of dimensionality. Using two-step semiparametric estimators can achieve semiparametric efficiency. However , as pointed out by Lee (2023), there is no result on the semiparametric efficiency of the two-step estimator in our context. Overall, the inference theory developed in this paper is easy to implement and useful for practitioners.

We also conduct a Monte Carlo simulation study to investigate the finite-sample performance of the PSMD estimator. The simulation results show that the PSMD estimator performs well in finite samples.

The rest of this paper is organized into sections. Section 2 introduces the model and briefly discusses the identification and PSMD estimation procedure. Section 3 develops the asymptotic theory for the PSMD estimator. The Monte Carlo simulation study is presented in Section 4, and Section 5 concludes this work. All mathematical proofs are provided in the Appendix.

We introduce several notations used throughout the paper. For a generic random variable A, the support of A is denoted by $Supp(A)$. For two random variables A and B, and for any $\tau \in (0, 1)$, $Q_{A|B}(\tau | b)$ indicates the τ-th conditional quantile of A on $B = b$, and $F_{A|B}(a | b)$ is the conditional distribution function of A given $B = b$. Furthermore, $\mathbb{E}[\cdot]$ is the expectation operator. For any positive real sequences $\{a_n\}$ and $\{b_n\}$, $a_n \lesssim b_n$ means that there exist finite constant $C > 0$ and $N \in \mathbb{N}$ such that $a_n \leq Cb_n$ for all $n \geq N$. If $a_n \lesssim b_n$ and $b_n \lesssim a_n$, it is denoted by $a_n \asymp b_n$.

II. Model and PSMD Estimation

A. Model and Identification

Recall the triangular model for QR in Lee (2022): for each $\tau \in (0, 1)$,

$$\begin{aligned} Y &= g(X, Z_1; \tau) + U(\tau), \\ X &= h(Z) + V, \end{aligned} \tag{1}$$

where $X \in \mathbb{R}^{d_x}$, $Z_1 \in \mathbb{R}^{d_{z1}}$, $Z \equiv (Z_1', Z_2')' \in \mathbb{R}^{d_{z1}+d_{z2}}$. In addition, $U(\tau)$ and V are unobserved error terms that are scalar, and Z_2 is a vector of excluded variables such that $Z_2 \in \mathbb{R}^{d_{z2}}$ and $d_{z_2} \geq d_x$. We call the first and second equations in Model (1) the outcome equation and the reduced-form equation, respectively.

To allow for the endogeneity of X, we assume that $U(\tau)$ and V can be correlated. The functions g and h are the parameters of interest that are nonparametrically specified, and researchers can only observe $(Y, X' Z')'$ from the data.

Suppose that $h(\cdot)$ is identified from the reduced-form equation. We further assume that

$$Q_{U(\tau)|Z,V}(\tau \mid Z, V) = Q_{U(\tau)|V}(\tau \mid V), \tag{2}$$

and that $Q_{U(\tau)|V}(\tau | V)$ is a function of V. Then, we have

$$Q_{U(\tau)|X,Z}(\tau \mid X, Z) = Q_{U(\tau)|X,V}(\tau \mid X, V) = Q_{U(\tau)|V}(\tau \mid V) \tag{3}$$

As a result, we have the following model restriction:

$$Q_{Y \mid X, Z, V}(\tau \mid X, Z, V) = g(X, Z_1 ; \tau) + r(V; \tau), \tag{4}$$

where $r(V;\tau) \equiv Q_{U(\tau)|V}(\tau|V)$ is an unknown function of V. The following set of conditions is from Lee (2022).

Assumption 2.1. The following conditions hold:

(i) There exists a known $(\bar{x}', \bar{z}_1')' \in Supp(X, Z_1)$ such that $g(\bar{x}, \bar{z}_1 ; \tau) = 0$;

(ii) $g(X, Z_1)$, $r(V)$, and $h(Z)$ are differentiable;

(iii) The boundary of $Supp(Z, V)$ has zero probability;

(iv) The function $h(\cdot)$ is identified over the support of Z;

(v) For each $\tau \in (0, 1)$, $Q_{U(\tau) \mid Z, V}(\tau \mid Z, V) = Q_{U(\tau) \mid V}(\tau \mid V)$ almost surely; and

(vi) $\Pr\left(rank\left(\frac{\partial h(Z)}{\partial Z_2}\right) = d_x \right) = 1$.

Notably, Assumption 2.1(iv) is a high-level but mild condition. Here, one can impose the conditional mean independence between Z and V, as in Newey *et al.* (1999), or the conditional quantile independence, as in Lee (2007). Assumption 2.1(vi) is a nonparametric rank condition similar to that in Newey *et al.* (1999) and is the key identifying assumption. For a detailed discussion on Assumption 2.1, one can refer to Lee (2022).

Proposition 2.1. (Theorem 2.1 in Lee (2022)) Suppose that Assumption 2.1 holds. Then, g, r, and h are nonparametrically identified over their support.

B. Penalized Sieve Minimum Distance Estimation

Let g_0, r_0, and h_0 be the true parameter values for g, r, and h, respectively. Let α denote the vector of parameters (*i.e.*, $\alpha \equiv (g, r, h)'$), and α_0 denote the true parameter vector.

We also denote the parameter spaces for g, r, and h by $\mathcal{G}$, $\mathcal{R}$, and $\mathcal{H}$, respectively. Then, we define the parameter space for α as the Cartesian product of $\mathcal{G}$, $\mathcal{R}$, and $\mathcal{H}$, and denote it by $\mathcal{A}$. We also let $\{W_i \equiv (Y_i, X_i', Z_i')' : i = 1, 2, \dots, n\}$ be the data and $\tau \in (0, 1)$ be given. Thus, we consider the following conditional-moment restriction:

$$m_\tau(X, Z ; \alpha) \equiv \mathbb{E}[\rho_\tau(W ; \alpha) \mid X, Z] = \mathbb{E}[1(Y \leq g(X, Z_1; \tau) + r(X - h(Z) ; \tau)) - \tau \mid X, Z], \tag{5}$$

where $1(\cdot)$ is an indicator function. Under the identification conditions, we have $m_\tau(X, Z ; \alpha) = 0$ almost certainly in X and Z if and only if $\alpha = \alpha_0$. For notational simplicity, we let $m(\cdot, \cdot ; \cdot) = m_\tau(\cdot, \cdot ; \cdot)$ for a given value of τ.

The PSMD sieve estimator of α_0, denoted by $\hat{\alpha}_n$, is defined as follows:

$$\hat{\alpha}_n \equiv \arg \inf_{\alpha \in A_n} \left\{ \frac{1}{n} \sum_{i}^{n} \hat{m}_n(X_i, Z_i ; \alpha)' \left[\hat{\Sigma}_n(X_i, Z_i) \right]^{-1} \hat{m}_n(X_i, Z_i ; \alpha) + \lambda_n \hat{P}_n(\alpha) \right\}, \tag{6}$$

where $\hat{m}_n(x, z ; \alpha)$ is a consistent estimator of $m(x, z ; \alpha)$, $\hat{\Sigma}_n(x, z)$ is a consistent estimator of the positive definite matrix $\Sigma(x, z)$, $\hat{P}_n(\alpha) \geq 0$ is a possibly random penalty function, and λ_n is a positive real sequence such that $\lambda_n \downarrow 0$. Furthermore, $\mathcal{A}_n$ is a sieve space for the parameter space $\mathcal{A}$.

To obtain the PSMD estimator $\hat{\alpha}_n$, we need to consider a nonparametric estimator of $m(X, Z ; \alpha)$. In this paper, we use the series estimation to consistently estimate $m(X, Z ; \alpha)$. Specifically, a series estimator $\hat{m}_n(X, Z ; \alpha)$ is given by

$$\hat{m}_n(X, Z ; \alpha) \equiv b^{J_n}(X, Z)'(B'B)^{-} \sum_{i=1}^{n} b^{J_n}(X_i, Z_i) \rho_\tau(W_i ; \alpha), \tag{7}$$

where $b_j(\cdot, \cdot)_{j=1}^{\infty}$ is a sequence of the basis functions,

$$b^{J_n}(x, z) \equiv \left(b_1(x, z), b_2(x, z), \ldots, b_{J_n}(x, z)\right)' \text{ and}$$

$$B \equiv \left[b^{J_n}(X_1, Z_1), b^{J_n}(X_2, Z_2), \ldots, b^{J_n}(X_n, Z_n)\right]'.$$

We also introduce a class of functions to define the parameter space. Let $f : \mathbb{D} \to \mathbb{R}$, where $\mathbb{D} \subseteq \mathbb{R}^{d_x}$ for some integer $d_x \geq 1$. Let $\omega = (\omega_1, \ldots, \omega_{d_x})$ be a d_x-tuple of nonnegative integers and define the differential operator as $\nabla^\omega f \equiv \frac{\partial^{|\omega|}}{\partial x_1^{\omega_1} \partial x_2^{\omega_2} \cdots \partial x_{d_x}^{\omega_{d_x}}} f(x)$, where $x = \left(x_1, x_2, \ldots, x_{d_x}\right) \in \mathbb{D}$ and $|\omega| \equiv \sum_{i=1}^{d_x} \omega_i$. We also let $[p]$ be the integer part *of* $p \in \mathbb{R}_+$ *and then a function* $f : \mathcal{X} \to \mathbb{R}$ *is* *called p-smooth if* it is $[p]$ times continuously differentiable on χ and for all ω such that $|\omega| = [p]$ and for some $\nu \in (0, 1]$ and constant $c > 0$, $\left|\nabla^{\omega} f(x) - \nabla^{\omega} f(y)\right| \leq c \cdot \|x - y\|_E^{\nu}$ for all $x, y \in \mathcal{X}$, where $\|\cdot\|_E$ is the Euclidean norm. Let $\mathcal{C}^{[p]}(\mathcal{X})$ denote the space of all $[p]$ times continuously differentiable real-valued functions on χ. A Hölder ball with smoothness p is defined below.

$$\Lambda_{\mathcal{C}}^{p}(\mathcal{X}) = \left\{ \begin{matrix} f \in \mathcal{C}^{[p]}(\mathcal{X}) : \sup_{|\omega| \leq [p]} \sup_{x \in \mathcal{X}} \left|\nabla^{\omega} f(x)\right| \leq C, \\ \sup_{|\omega| = [p]} \sup_{x, y \in \mathcal{X}, x \neq y} \dfrac{\left|\nabla^{\omega} f(x) - \nabla^{\omega} f(y)\right|}{\|x - y\|_E^{\nu}} \leq C \end{matrix} \right\}$$

In the equation above, C is a positive finite constant.

For a generic function defined on the support of a random variable X with its distribution function F_x, χ, let $\|f\|_{\infty} \equiv ess \sup_{x \in \mathcal{X}} |f(x)|$ and $\|f\|_2^2 \equiv \int f(x)^2 \, dF_X(x)$ denote the supremum-norm (or sup-norm) and L_2-norm, respectively, while ess sup denotes the essential supremum. For any $\alpha, \tilde{\alpha} \in \mathcal{A}$, define $\|\alpha - \tilde{\alpha}\|_{\mathcal{A}, \infty} \equiv \|g(\cdot, \cdot) - \tilde{g}(\cdot, \cdot)\|_{\infty} + \|r(\cdot) - \tilde{r}(\cdot)\|_{\infty} + \left\|h(\cdot) - \tilde{h}(\cdot)\right\|_{\infty}$ and $\|\alpha - \tilde{\alpha}\|_{\mathcal{A}, 2}^2 \equiv \|g(\cdot, \cdot) - \tilde{g}(\cdot, \cdot)\|_2^2 + \|r(\cdot) - \tilde{r}(\cdot)\|_2^2 + \left\|h(\cdot) - \tilde{h}(\cdot)\right\|_2^2$. For a (random) vector A, $\|A\|_E$ is the Euclidean norm of A.

III. Asymptotic Theory

A. Consistency and Convergence Rate

We impose the conditions below to establish the consistency of the PSMD estimator.

Assumption 3.1

(i) The data $\{W_i : i = 1, 2, \ldots, n\}$ are i.i.d; (ii) The conditional distribution of Y on X and Z admits its conditional density function $f_{Y|X,Z}$ such that $f_{Y|X,Z}(g_0(X, Z_1; \tau) + r_0(X - h_0(Z); \tau) \mid X, Z) > 0$ almost certainly, $f_{Y|X,Z}(y|x, z)$ is continuous in $(y, x', z')'$ and $\sup_y f_{Y|X,Z}(y \mid x, z) < \infty \in Supp(X, Z)$; (iii) $Supp(X, Z)$ is a compact subset of $\mathbb{R}^{d_x + d_z}$ with Lipschitz continuous boundary; and (iv) the density function of (X, Z), $f_{XZ}(x, z)$, is bounded and bounded away from zero over $Supp(X, Z)$.

Assumption 3.2. (i) $g_0 \in \mathcal{G} \equiv \Lambda_{c_g}^{p_g}(Supp\,(X,\ Z_1))$, $r_0 \in \mathcal{R} \equiv \Lambda_{c_r}^{p_r}(Supp\,(X - h_0\,(Z)))$, and $h_0 \in \mathcal{H} \equiv \Lambda_{c_h}^{p_h}\left(Supp\,(Z)\right)$ with $p_g > d_x + d_{z_1}$, $p_r > d_x$, and $p_h > d_z$; and (ii) all first-order partial derivatives of g_0, r_0, and h_0 are uniformly bounded.

Assumption 3.3. (i) For a sequence of basis functions, $\left(p_j\,(\cdot)\right)_{j=1}^{\infty}$, the sieve spaces for $\mathcal{G}$, $\mathcal{R}$, and $\mathcal{H}$ are given by

$$\mathcal{G}_n \equiv \{g_n\,(x,\ z_1) = p^{k_{g,n}}\,(x,\ z_1)'\beta_{g,n} : \|g_n\|_\infty \leq c_g\}$$
$$\mathcal{R}_n \equiv \{r_n\,(v) = p^{k_{r,n}}\,(v)'\beta_{r,n} : \|r_n\|_\infty \leq c_r\}$$
$$\mathcal{H}_n \equiv \{h_n\,(z) = p^{k_{h,n}}\,(z)'\beta_{h,n} : \|h_n\|_\infty \leq c_h\}$$

where $k_{g,n}$, $k_{r,n}$, *and* $k_{h,n}$ are some positive nondecreasing integer sequences such that $k_{g,n}$, $k_{r,n}$, $k_{h,n} \to \infty$, $\max(k_{g,n},\ k_{r,n},\ k_{h,n}) = o(n)$; (ii) let $\mathbb{Q}_{g,n} \equiv \mathbb{E}\left[p^{k_{g,n}}(X,\ Z_1)\cdot p^{k_{g,n}}(X,\ Z_1)'\right]$, $\mathbb{Q}_{r,n} \equiv \mathbb{E}\left[p^{k_{r,n}}(V)p^{k_{r,n}}\,(V)'\right]$, and $\mathbb{Q}_{h,n} \equiv \mathbb{E}\left[p^{k_{h,n}}\,(Z)p^{k_{h,n}}\,(Z)'\right]$, then the eigenvalues of $\mathbb{Q}_{g,n}$, $\mathbb{Q}_{r,n}$, and $\mathbb{Q}_{h,n}$ are bounded above and away from zero uniformly over all n; and (iii) there exist $\{\pi_n g_0\}_n$, $\{\pi_n r_0\}_n$, and $\{\pi_n h_0\}_n$ such that $\|g_0 - \pi_n g_0\|_\infty = O\left(k_{g,n}^{-\sigma_g}\right)$, $\|r_0 - \pi_n r_0\|_\infty = O\left(k_{r,n}^{-\sigma_r}\right)$, and $\|h_0 - \pi_n h_0\|_\infty = O\left(k_{h,n}^{-\sigma_h}\right)$ for some σ_g, σ_r, $\sigma_h > 0$.

Assumption 3.4. (i) $\Sigma\,(X,\ Z) = \widehat{\Sigma}_n\,(X,\ Z) = \tau(1-\tau)$ almost surely for all n; (ii) $\lambda_n > 0$ for all n, $\lambda_n = o(n^{-1})$, and $\|\alpha_0 - \pi_n\alpha_0\|_{\mathcal{A},\,\infty} = O\,(\lambda_n)$; (iii) $P\,(\alpha) = \hat{P}_n\,(\alpha) = \|\nabla g\|_2^2$.

Assumption 3.5.
$F_{X\mid Y,\,Z}\left(g_0\,(\cdot,\ \cdot\ ;\ \tau) + r_0\,(\cdot - h_0\,(\cdot\ ;\ \tau)\mid X = \cdot,\ Z = \cdot)\right) \in \Lambda_{c_m}^{p_m}\,(Supp\,(X,\ Z))$ with $p_m > 1/2$.

Assumption 3.6. The following conditions hold:

(i) $\left(b_j\,(\ \cdot\ )\right)_{j=1}^{\infty}$ is a sequence of polynomial spline function;

(ii) $\max_{j\leq J_n}\mathbb{E}\left[\|b_j\,(X,\ Z)\|_E^2\right] < C < \infty$ for some constant C;

(iii) the smallest eigenvalue $\mathbb{E}\left[b^{J_n}\,(X,\ Z)b^{J_n}\,(X,\ Z)'\right]$ is bounded away from zero for all J_n; and

(iv) $J_n \to \infty$ as $n \to \infty$, and $J_n\log(J_n) = o(n)$.

Assumption 3.1 is standard in the literature on QR and sieve estimation, Assumption 3.2 defines the parameter space for each structural parameter, and Assumption 3.3 defines the sieve spaces for $\mathcal{A}$. Condition (iii) of Assumption 3.3 requires that the sieve spaces approximate the structural functions well in terms of the approximation error shrinking at a certain rate. Under Assumptions 3.1 and 3.2, various sieve spaces can satisfy this condition. Specifically, for $f \in \Lambda_c^p(\mathcal{X})$, with χ being a compact subset of $\mathbb{R}^d$, we have $\|f - \pi_n f\|_\infty = O\left(k_n^{-\frac{p}{d}}\right)$ by choosing polynomial or spline sieve spaces (*cf.* Newey, 1997). Assumption 3.4 defines the weighting matrix and the penalty function. Note that the weighting matrix in the assumption is the variance of the moment condition in (5). We consider the L_2-norm of the derivative of g as the penalty function, which is widely used in the literature. Assumption 3.5 implies that the conditional moment function $m(X, Z; \alpha_0)$ belongs to a Hölder ball, and Assumption 3.6 specifies the sieve space on which $m(\cdot,\cdot)$ is approximated.

Theorem 3.1. Suppose that Assumptions 2.1 and 3.1–3.6 hold. Then,

$$\|\hat{\alpha}_n - \alpha_0\|_{\mathcal{A},\infty} = o_p(1).$$

Given the consistency result in Theorem 3.1, we can restrict our attention to a $\|\cdot\|_{\mathcal{A},\infty}$ - shrinking neighborhood of α_0 to establish the convergence rate of $\hat{\alpha}_n$. For given small $\varepsilon > 0$ and large $M > 0$, we define

$$\mathcal{A}_{os} \equiv \{\alpha \in \mathcal{A} : \|\alpha - \alpha_0\|_{\mathcal{A},\infty} \leq \varepsilon,\ \|\alpha\|_{\mathcal{A},\infty} \leq M\},$$
$$\mathcal{A}_{osn} \equiv \mathcal{A}_{os} \bigcap \mathcal{A}_n.$$

We also define

$$\frac{dm(X, Z; \alpha_0)}{d\alpha}[\alpha - \alpha_0] \equiv \left.\frac{d\mathbb{E}[\rho(W; (1-t)\alpha_0 + t\alpha) \mid X, Z]}{dt}\right|_{t=0}$$

as the pathwise derivative of m in the direction $[\alpha - \alpha_0]$ evaluated at α_0. Let $\|\cdot\|$ denote a pseudo metric on $\mathcal{A}_{os}$, where we have

$$\|\alpha_1 - \alpha_2\| \equiv \sqrt{\mathbb{E}\left[\left(\frac{dm(X,Z;\ \alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)' \left(\sum(X,Z)\right)^{-1}\left(\frac{dm(X,Z;\ \alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)\right]}.$$

for any $\alpha_1,\ \alpha_2 \in \mathcal{A}_{os}$.

Assumption 3.7.
(i) $\mathcal{A}_{os}$ and $\mathcal{A}_{osn}$ are convex; and (ii) $\mathbb{E}\left[\|m(X,\ Z\ ;\ \alpha)\|_E^2\right] \asymp \|\alpha - \alpha_0\|^2$ for all $\alpha \in \mathcal{A}_{osn}$.

Assumption 3.8.

$$\max\left\{\frac{J_n}{n},\ J_n^{-\frac{2p_m}{d_x+d_z}},\ \|\alpha_0 - \pi_n\alpha_0\|_{\mathcal{A},\infty},\ \lambda_n,\ o(n^{-1})\right\} = \max\left\{\frac{J_n}{n},\ J_n^{-\frac{2p_m}{d_x+d_z}}\right\},$$

where $\|\alpha_0\ -\ \pi_n\alpha_0\|_{\mathcal{A},\infty} = \max\left\{k_{g,n}^{-\sigma_g},\ k_{r,n}^{-\sigma_r},\ k_{h,n}^{-\sigma_h}\right\}$.

Assumption 3.7 is mild, as we restrict the attention to $\alpha \in \mathcal{A}_{osn}$ (*cf.* Van de Geer, 2000). Assumption 3.8 further restricts the rate at which λ_n approaches zero. The following theorem establishes the L_2-convergence rate of the PSMD estimator $\hat{\alpha}_n$.

Theorem 3.2. Suppose that Assumptions 2.1 and 3.1–3.8 hold. Then, for $\|\alpha_0\ -\ \pi_n\alpha_0\|_{\mathcal{A},\infty} = \max\left\{k_{g,n}^{-\sigma_g},\ k_{r,n}^{-\sigma_r},\ k_{h,n}^{-\sigma_h}\right\}$, we have

$$\|\hat{\alpha}_n - \alpha_0\|_{\mathcal{A},2} = O_p\left(\max\left\{\|\alpha_0 - \pi_n\alpha_0\|_{\mathcal{A},\infty},\ \sqrt{\frac{J_n}{n}},\ J_n^{-\frac{p_m}{d_x+d_z}}\right\}\right).$$

If $J_n \asymp k_n = \max\left(k_{g,n},\ k_{r,n},\ k_{h,n}\right)$, then

$$\|\hat{\alpha}_n - \alpha_0\|_{\mathcal{A},2} = O_p\left(\max\left\{\|\alpha_0\ -\ \pi_n\alpha_0\|_{\mathcal{A},\infty},\ \sqrt{\frac{k_n}{n}},\ k_n^{-\frac{p_m}{d_x+d_z}}\right\}\right)$$

The convergence rates presented in Theorem 3.2 are standard nonparametric convergence rates (*cf.* Stone, 1982), thus implying that the sieve estimator does not suffer from an ill-posed inverse problem.

Although the structural function g_0 depends on the endogenous regressor X, we can effectively circumvent the ill-posed inverse problem by virtue of the triangular system of equations.

B. Sieve Wald and QLR Inference

We now consider inference on functionals of the parameter α. Here, we use the novel approach developed by Chen and Pouzo (2015), who considered sieve Wald and QLR test statistics for inference on the functionals. One important advantage of their approach is that the verification of whether the functional of interest is regular or irregular is not required. We denote the L_2-convergence rate of the PSMD estimator provided in Theorem 3.2 by $\delta^*_{2,n}$; that is,

$$\delta^*_{2,n} = \max\left\{k_{g,n}^{-\sigma_g},\ k_{r,n}^{-\sigma_r},\ k_{h,n}^{-\sigma_h},\ \sqrt{\frac{J_n}{n}},\ J_n^{-\frac{p_m}{d_x + d_z}}\right\}.$$

We let $\delta_{2,n} \equiv \log(\log(n+1)) \cdot \delta^*_{2,n}$ and assume that $\delta_{2,n} = o(1)$.
Define

$$\mathcal{N}_{os} \equiv \{\alpha \in \mathcal{A} : \|\alpha - \alpha_0\| \le \delta_{2,n}\},$$
$$\mathcal{N}_{osn} \equiv \mathcal{N}_{os} \bigcap \mathcal{A}_n$$

Let $f : \mathcal{A} \to \mathbb{R}$ be a functional defined over the parameter space. For any $v \in \mathcal{A}$, define

$$\frac{df(\alpha_0)}{d\alpha}[v] \equiv \left.\frac{\partial f(\alpha_0 + tv)}{\partial t}\right|_{t=0}$$

as a pathwise derivative of the functional f at α_0 in the direction of v. We assume that $\frac{df(\alpha_0)}{d\alpha}[\,\cdot\,] : \mathcal{A} \to \mathbb{R}$ is a bounded linear functional with respect to $\|\cdot\|_{\mathcal{A},2}$.
Many functionals of interest pertain to empirical analysis. For example, one can consider the (weighted) average of the derivative of g_0 (*i.e.*, $f(\alpha_0) = \int \nabla_x g(x, z_1)\, dF_{X,Z_1}(x, z_1)$). In addition, should one be interested in the function value at some point, the functional can be defined as $f(\alpha_0) = g_0(\bar{x}, \bar{z}_1)$ for some $(\bar{x}', \bar{z}_1')' \in Supp(X, Z_1)$.[2]

[2] Although we focus on nonparametric models, when the model contains a

Let $\mathbb{V}$ be a linear span of $\mathcal{A}_{os} - \{\alpha_0\}$ endowed with both $\| \cdot \|_2$ and $\| \cdot \|$ norms. Note that under Assumption 3.1(ii), there exists a finite constant $C > 0$ such that $\| v \| \leq C \cdot \| v \|_2$ for any $v \in \mathbb{V}$. Let $\bar{\mathbb{V}}$ be the closure of $\mathbb{V}$ with respect to $\| \cdot \|$. For any $v_1, v_2 \in \bar{\mathbb{V}}$, define an inner product induced by the metric $\| \cdot \|$ as follows:

$$< v_1, v_2 > \equiv \mathbb{E}\left[\left(\frac{dm(X, Z ; \alpha_0)}{d\alpha}[v_1]\right)' \sum (X, Z)^{-1} \left(\frac{dm(X, Z ; \alpha_0)}{d\alpha}[v_2]\right)\right], \quad (8)$$

where

$$\frac{dm(X, Z ; \alpha_0)}{d\alpha}[u_n^*] = f_{Y \mid X, Z}\left(g_0 (X, Z_1) + r_0 (X - h_0 (Z)) \mid X, Z\right) \cdot \left[u_{g,n}^* + u_{r,n}^* + \nabla r_0 \cdot u_{h,n}^*\right]. \quad (9)$$

The inner product is well-defined under Assumption 3.1, and $\left(\bar{\mathbb{V}}_n, \| \cdot \|\right)$ is a Hilbert space. Let $\alpha_{0,n} \in \mathcal{A}_n$ be the projection of α_0 onto $\mathcal{A}_n$ under $\| \cdot \|$; that is,

$$\left\|\alpha_{0,n} - \alpha_0\right\| = \arg \min_{\alpha \in \mathcal{A}_n} \|\alpha - \alpha_0\|$$

Furthermore, let $\bar{\mathbb{V}}_n$ be the closed linear span of $\mathcal{A}_{osn} - \{\alpha_0\}$ under $\| \cdot \|$. Then, $\left(\bar{\mathbb{V}}_n, \| \cdot \|\right)$ is a finite dimensional Hilbert space. Given that every linear functional on a finite dimensional Hilbert space is bounded, the Riesz representation theorem implies that $v_n^* \in \bar{\mathbb{V}}_n$ exists such that, for any $v \in \bar{\mathbb{V}}_n$, we have

$$\frac{df(\alpha_0)}{d\alpha}[v] = < v_n^*, v > \quad (10)$$

and

$$\left\|v_n^*\right\| \equiv \sup_{v \in \bar{\mathbb{V}}_n : \|v\| \neq 0} \frac{\left|\frac{df(\alpha_0)}{d\alpha}[v]\right|}{\|v\|} < \infty \quad (11)$$

finite-dimensional parameter $\theta_0 \in \mathbb{R}^{d_\theta}$, we can set $f(\alpha_0) = \lambda'\theta_0$ for some known $\lambda \in \mathbb{R}^{d_\theta}$ to perform statistical inference on θ_0.

where v_n^* is called the sieve Riesz representer of the functional $\frac{df(\alpha_0)}{d\alpha}[\cdot]$ on $\bar{\mathbb{V}}_n$. For each n, define the sieve score associated with the i-th observation as

$$S_{n,i}^* \equiv \left(\frac{dm(X_i, Z_i; \alpha_0)}{d\alpha}[v_n^*]\right)' \sum(X_i, Z_i)^{-1} \rho(W_i; \alpha_0)$$

and let

$$\left\|v_n^*\right\|_{sd} \equiv Var(S_{n,i}^*)$$

to denote the sieve variance. We also defined the scaled sieve Riesz representer as

$$u_n^* \equiv \frac{v_n^*}{\left\|v_n^*\right\|_{sd}}.$$

We impose the following condition on the functional of interest.

Assumption 3.9. Let $\mathcal{J}_n \equiv \{t \in \mathbb{R} : |t| \leq 4M_n\delta_{2,n}\}$. Then, the following conditions hold:

(i) $v \mapsto \frac{df(\alpha_0)}{d\alpha}[v]$ is a nonzero linear functional mapping from $\mathbb{V}$ to $\mathbb{R}$, $\left\{\bar{\mathbb{V}}_n\right\}_{n=1}^{\infty}$ is an increasing sequence of finite dimensional Hilbert space that is dense in $\left(\bar{\mathbb{V}}, \|\cdot\|\right)$, and $\frac{\left\|v_n^*\right\|}{\sqrt{n}} = o(1)$;

(ii)

$$\sup_{(\alpha, t) \in \mathcal{N}_{osn} \times \mathcal{J}_n} \frac{\sqrt{n}\left|f(\alpha + tu_n^*) - f(\alpha_0) - \frac{df(\alpha_0)}{d\alpha}[\alpha + tu_n^* - \alpha_0]\right|}{\left\|v_n^*\right\|} = o(1);$$

(iii)

$$\frac{\sqrt{n}\left|\frac{df(\alpha_0)}{d\alpha}[\alpha_{0,n} - \alpha_0]\right|}{\left\|v_n^*\right\|} = o(1).$$

Condition (i) of Assumption 3.9 essentially places a restriction on the rate at which the sieve space grows as n goes to the infinity. Meanwhile, Condition (ii) of Assumption 3.9 restricts the bias caused by the nonlinearity of $f(\cdot)$. This condition requires that $f(a) - f(a_0)$ for α in a neighborhood of α_0 be well-approximated by the pathwise derivative of f.

This condition is satisfied if $f(\cdot)$ is a linear functional, such as weighted average derivative and the structural function value at a given point. Condition (iii) of Assumption 3.9 is required to eliminate the asymptotic bias caused by the finite-dimensional sieve approximation of α_0, $\alpha_{0,n}$.

Assumption 3.10. (i) $\frac{dm(X, Z ; \alpha_0)}{d\alpha}[u_n^*] \in \Lambda_{c_A}^{p_A}(Supp(X, Z))$ with $p_A > \frac{(d_x + d_z)}{2}$ and $c_A > 0$; (ii) let $f'_{Y \mid X, Z}(y \mid x, z) \equiv \frac{\partial f_{Y \mid X, Z}(y \mid x, z)}{\partial y}$ denote the derivative of the conditional density function of Y given $X = x$ and $Z = z$. The function $f'_{Y \mid X, Z}(\cdot \mid x, z)$ is continuous and uniformly bounded for almost all $(x', z')' \in Supp(X, Z)$.

Assumption 3.11. The following conditions hold:

(i) $\delta_{2,n}^* = \max\left\{k_{g,n}^{-\sigma_g}, k_{r,n}^{-\sigma_r}, k_{h,n}^{-\sigma_h}, \sqrt{\frac{J_n}{n}}, J_n^{-\frac{p_m}{d_x + d_z}}\right\} = \sqrt{\frac{J_n}{n}}$;

(ii) $\left(\log\left(\log(1+n)\right)\right)^8 J_n^5 \cdot \max\{k_{g,n}, k_{r,n}, k_{h,n}\}^4 = o(n)$;

(iii) $J_n^{-\frac{2p_a}{(d_x + d_z)}} \delta_{2,n}^2 = o(n^{-1})$.

Assumption 3.12. There exists $\delta_0 > 0$ such that $\mathbb{E}\left[\left|\frac{s_{n,i}^*}{\|v_n^*\|_{sd}}\right|^{2+\delta_0}\right] < \infty$.

Assumption 3.10 imposes a smoothness condition on the derivative of m and the conditional density function of Y given X and Z. Assumption 3.11 restricts the rates at which $k_{g,n}$, $k_{r,n}$, $k_{h,n}$, and J_n grow. This assumption is required to control for the bias term of the sieve estimator. Assumption 3.12 is a sufficient condition for the Linderberg's condition that allows to employ the central limit theorem for the score function. The following theorem establishes the asymptotic normality of the sieve plug-in estimator of the functional.

Theorem 3.3. Suppose that Assumptions 2.1 and 3.1-3.12 hold. Then,

$$\frac{\sqrt{n}\left(f(\hat{\alpha}_n) - f(\alpha_0)\right)}{\|v_n^*\|_{sd}} \xrightarrow{d} N(0, 1).$$

Since the generalized residual function $\rho(W ; \alpha)$ in our case is non-smooth in α, estimating the sieve Riesz representer may be cumbersome, as pointed out by Chen and Pouzo (2015). For this reason, we propose the sieve QLR statistic for inference on the functional of interest. To this end, we define the restricted sieve space under $H_0 : f(a_0) = f_0$ for some known $f_0 \in \mathbb{R}$ as

$$\mathcal{A}_n^R(f_0) \equiv \{\alpha \in \mathcal{A}_n : f(\alpha) = f_0\}$$

Let $\hat{a}_n^R$ denote the restricted approximate PSMD estimator, defined as

$$\hat{Q}_n(\hat{\alpha}_n^R) + \lambda_n \hat{P}_n(\hat{\alpha}_n^R) \le \inf_{\alpha \in \mathrm{A}_n^R(f_0)} \left\{\hat{Q}_n(\alpha) + \lambda_n \hat{P}_n(\alpha)\right\} + o_p(n^{-1})$$

where $\hat{Q}_n(\alpha) = \dfrac{1}{n}\sum_i^n \hat{m}_n(X_i, Z_i; \alpha)' \left[\hat{\Sigma}_n(X_i, Z_i)\right]^{-1} \hat{m}_n(X_i, Z_i; \alpha)$.

Then, the sieve QLR statistic is defined as

$$QLR_n(f_0) \equiv n\left(\hat{Q}_n(\hat{\alpha}_n^R) - \hat{Q}_n(\hat{\alpha}_n)\right).$$

The following theorem establishes the asymptotic distribution of $\widehat{QLR}_n(f_0)$:

Theorem 3.4. Suppose that Assumptions 2.1 and 3.1-3.12. Then, under $H_0 : f(a_0) = f_0$, we have

$$\widehat{QLR}_n(f_0) \xrightarrow{\quad d \quad} \chi^2(1).$$

Theorem 3.4 provides a way to construct confidence sets for the functional $f(a_0)$. Let c_p denote the $p \times 100\%$ quantile of $\chi^2(1)$ for given $p \in (0, 1)$ and define $\hat{Q}_n^*(f_0) \equiv \inf_{\alpha \in \mathcal{A}_n^R(f_0)} \hat{Q}_n(\alpha)$. Then, a $(1 - p) \times 100\%$ confidence region for $f(a_0)$ can be constructed as

$$CS_{1-p} = \{c \in \mathbb{R} : \widehat{QLR}_n(c) \le c_{1-p}\} \tag{12}$$

Notably, Theorems 3.3 and 3.4 are applicable, regardless of whether or not $f(a_0)$ is $\sqrt{n}$-estimable. Based on Theorem 3.3, we can see in a straightforward manner that the asymptotic variance of $\sqrt{n}\left(f(\hat{\alpha}_n) - f(\alpha_0)\right)$ is given by $\left\|\upsilon_n^*\right\|$. Here, $\left\|\upsilon_n^*\right\|$ is allowed to diverge to the infinity, which is the case that $f(a_0)$ is not $\sqrt{n}$-estimable (*cf.* Lemma 3.3 in Chen and Pouzo (2015)).

IV. Monte Carlo Simulation

Next, we conduct a Monte Carlo simulation study to investigate the finite-sample performance of the PSMD estimator. Previously, Lee (2022) has shown that the PSMD estimator without the penalty performs well in the finite sample. Here, we consider the case in which a nontrivial penalty function is incorporated. The data generating process (DGP) is as follows:

$$\begin{aligned} Y &= 2\Big(F_B\left(X/2\ ;\ a_g(\tau),\ b_g(\tau)\right) - F_B\left(0.5\ ;\ a_g(\tau),\ b_g(\tau)\right)\Big) + U \\ X &= F_B\left(Z/2 + 0.5\ ;\ a_h,\ b_h\right) + V \end{aligned},$$

where $U = \Phi(\varepsilon_1) - 0.5$ and $V = \Phi(\varepsilon_2) - 0.5$ with $(\varepsilon_1,\ \varepsilon_2)' \sim BVN\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0.3\\0.3 & 1\end{bmatrix}\right)$. BVN stands for the bivariate normal distribution, $\Phi(\cdot)$ is the standard normal distribution function, and $F_B(\cdot\,;\ a,\ b)$ is the beta distribution function with parameters a and b.

For the quantile level of interest, we consider 0.25, 0.5, and 0.75 quantile levels (*i.e.*, $\tau \in \{0.25.\ 0.5,\ 0.75\}$) and allow for $a_g(\tau)$ and $b_g(\tau)$ to vary across the quantile level τ by setting $a_g(\tau) = 4 + \Phi^{-1}(\tau)$ and $b_g(\tau) = 4 - \Phi^{-1}(\tau)$. For the reduced-form equation, we set $a_h = b_h = 2$. Note that the normalization value $\bar{x}$ in Assumption 2.1 is 1 in this simulation. The sample size n is set to be 500.

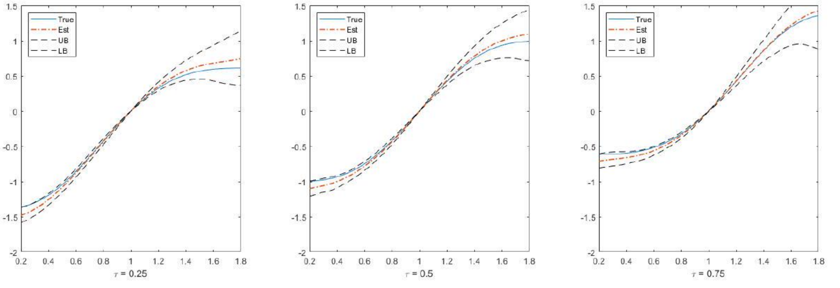
The sieve spaces for $\mathcal{G}$ and $\mathcal{H}$ are chosen to be polynomial sieves, and Hermite polynomial sieve is used for $\mathcal{R}$ with $k \propto n^{1/7}$, where $k_n = \max(k_{g,n},\ k_{r,n},\ k_{h,n})$. We use a series estimator of $m_\tau(\cdot\,,\cdot\,;\alpha)$ by using the tensor product of two polynomial spline sieve spaces as m_τ is a function of X and Z. In our simulation, we set $k_n = 5$ by selecting a proper constant C_0 such that $k_n = C_0 n^{1/7} = 5$ and $J_n = 17$.

We consider the $IBIAS^2$ and $IVAR$ of the PSMD estimator of $g_0(\cdot\,;\tau)$ as finite-sample performance measures of the PSMD estimator. Both $IBIAS^2$ and $IVAR$ are computed via numerical integration over the [0.2, 1.8] with a grid size of 0.01. The integrated mean squared error (MSE) is defined as the sum of $IBIAS^2$ and $IVAR$. We employ the (empirical) L_2 norm of ∇g over the [0.2, 1.8] for the penalty function, and tuning parameter for penalization, λ_n, varies from 0 to 0.0001.[3] All results are

[3] The support of X is [0, 2], and the integration region is an interior of the support.

TABLE 1
SIMULATION RESULTS (n = 500, λ_n = 1×10^{-5})

	$\hat{g}_n$ (·; 0.25)	$\hat{g}_n$ (·; 0.5)	$\hat{g}_n$ (·; 0.75)
IBIAS^2	0.0037	0.0026	0.0016
IVAR	0.0033	0.0033	0.0034
IMSE	0.0070	0.0059	0.0051



Note: The solid lines are the true structural function g_0, and the dash-dotted lines are the (pointwise average of) sieve estimator $\hat{g}_n$. We report 99% Monte Carlo confidence bands which are dashed lines in each figure.

FIGURE 1
SIMULATION RESULTS (n = 500, λ_n = 1×10^{-5})

obtained from 1,000 simulation iterations.

Table 1 shows the simulation results for each $\tau \in \{0.25. 0.5, 0.75\}$ with λ_n = 1×10^{-5}. Regardless of which quantile levels are used, we can see that the PSMD estimator performs well in finite samples in terms of both *IBIAS*^{2} and *IVAR*. Figure 1 shows the corresponding estimated curves and Monte Carlo confidence bands. In the figure, the true function is depicted as a solid line; the corresponding estimator is represented by a dotted line, which constitutes the pointwise average derived from 1000 Monte Carlo simulations; and the Monte Carlo 99% confidence bands are illustrated with dashed lines.

We report the simulation results for $\hat{g}_n(\cdot)$ with various values of λ_n in Table 2 to investigate the sensitivity of the PSMD estimator to λ_n. Following Chen and Pouzo (2012), we consider the "sieve dominating case," in which the role of penalty is relatively small (*i.e.*, $\lambda_n \searrow 0$ fast enough).[4] The results in Table 2 indicate that the PSMD estimators are

[4] Chen and Pouzo (2012) also considered the class of PSMD estimators using

TABLE 2
SENSITIVITY TO λ_n(n = 500, τ = 0.5)

λ_n	1×10^{-4}	7×10^{-5}	5×10^{-5}	3×10^{-5}	1×10^{-5}	0
$IBIAS^2$	0.0185	0.0162	0.0075	0.0018	0.0026	0.0027
IVAR	0.0028	0.0035	0.0066	0.0047	0.0033	0.0035
IMSE	0.0213	0.0197	0.0141	0.0065	0.0059	0.0062

not sensitive to the choice of λ_n. In addition, the *IMSE* tends to decrease as the degree of penalization decreases (*i.e.*, $\lambda_n \downarrow 0$).

At this point, the performance of the test statistics should be considered. However, as Chen and Pouzo (2015) already investigated the finite sample performance of the test statistics in their simulation study, we no longer conducted a simulation study in this paper.

V. Conclusion

In this paper, we consider the nonparametric estimation and inference for QR with endogenous regressors. The identification of the model parameter is achieved using a control function approach. Based on the identification result, we propose to use the PSMD estimation procedure, which is practical, useful, and easy to implement . We also develop the asymptotic theory for the PSMD estimator, including consistency, convergence rates, asymptotic normality, and the distributional theory for the sieve QLR test statistics. The results of the Monte Carlo simulation study confirm that the PSMD estimator performs well in finite samples.

(Submitted Jan 15, 2025; Accepted Jan 15, 2025)

large lower semicompact penalty. The role of penalization in our context is not substantial given that the nonparametric objects in the estimating equation do not depend on the endogenous regressors once we include the control function. Consequently, we use a small penalty in the simulation study.

References

Ackerberg, D., Chen, X., and Hahn, J., "A practical asymptotic variance estimator for two-step semiparametric estimators." *Review of Economics and Statistics 94*(No. 2 2012): 481-498.

Bitler, M. P., Gelbach, J. B., and Hoynes, H. W., "What mean impacts miss: Distributional effects of welfare reform experiments." *American Economic Review 96*(No. 4 2006): 988-1012.

Bitler, M. P., Gelbach, J. B., and Hoynes, H. W., "Distributional impacts of the self-sufficiency project." *Journal of Public Economics 92*(No. 3-4 2008): 748-765.

Chen, X., "Large sample sieve estimation of semi-nonparametric models." In Heckman, J. J., and Leamer, E. E. (eds.): *Handbook of Econometrics 6B*, 2007: 5549-5632.

Chen, X. and Pouzo, D., "Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals." *Econometrica 80*(No. 1 2012): 277-321.

Chen, X. and Pouzo, D., "Sieve Wald and QLR inferences on semi/nonparametric conditional moment models." *Econometrica 83*(No. 3 2015): 1013-1079.

Chernozhukov, V., Fernández-Val, I., and Kowalski, A. E., "Quantile regression with censoring and endogeneity." *Journal of Econometrics 186*(No. 1 2015): 201-221.

Chernozhukov, V. and Hansen, C., "Quantile models with endogeneity." *Annual Review of Economics 5*(No. 1 2013): 57-81.

Das, M., Newey, W. K., and Vella, F., "Nonparametric estimation of sample selection models." *The Review of Economic Studies 70*(No. 1 2003): 33-58.

Hahn, J., Liao, Z., and Ridder, G., "Nonparametric two-step sieve M estimation and inference." *Econometric Theory 34*(No. 6 2018): 1281-1324.

Hahn, J., and Ridder, G., "Asymptotic variance of semiparametric estimators with generated regressors." *Econometrica 81*(No. 1 2013): 315-340.

Koenker, R. and Bassett, G., "Regression quantiles." *Econometrica 46*(No. 1 1978): 33-50.

Kosorok, M. R., *Introduction to empirical processes and semiparametric inference*, Springer, 2008.

Lee, S., "Endogeneity in quantile regression models: A control function

approach." *Journal of Econometrics 141*(No. 2 2007): 1131-1158.
Lee, S., "Nonparametric estimation of a triangular system of equations for quantile regression." *Journal of Economic Theory and Econometrics 33*(No. 4 2022): 31-53.
Lee, S., "Efficient estimation of a triangular system of equations for quantile regression." *Economics Letters 226*(2023): 111085.
Mammen, E., Rothe, C., and Schienle, M., "Nonparametric regression with nonparametrically generated covariates." *The Annals of Statistics 40*(No. 2 2012): 1132-1170.
Newey, W. K., "Convergence rates and asymptotic normality for series estimators." *Journal of Econometrics 79*(No. 1 1997): 147-168.
Newey, W. K., "Two-step series estimation of sample selection models." *The Econometrics Journal 12*(No. S1 2009): S217-S229.
Newey, W. K., Powell, J. L., and Vella, F., "Nonparametric estimation of triangular simultaneous equations models." *Econometrica 67*(No. 3 1999): 565-603.
Stone, C. J., "Optimal global rates of convergence for nonparametric regression." *The Annals of Statistics 10*(No. 4 1982): 1040-1053.
Van de Geer, S. A., *Empirical processes in M-estimation*, Cambridge University Press, 2000.

Appendix

A. Mathematical Proofs

In this section, we provide mathematical proofs of the main results. We introduce notation that will be used in the proofs. Let $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ be a metric space of real valued function $f : \mathcal{X} \to \mathbb{R}$. The covering number $N(\varepsilon, \mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is the minimum number of $\|\cdot\|_{\mathcal{F}}$ ε balls that cover $\mathcal{F}$. The entropy is the logarithm of the covering number. An ε-bracket in $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is a pair of functions $l, u \in \mathcal{F}$ such that $\|l\|_{\mathcal{F}}, \|u\|_{\mathcal{F}} < \infty$ and $\|u - l\|_{\mathcal{F}} \leq \varepsilon$. The covering number with bracketing $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is the minimum number of $\|\cdot\|_{\mathcal{F}}$ ε-brackets that cover $\mathcal{F}$. The entropy with bracketing is the logarithm of the covering number with bracketing. The bracketing integral is defined as $\int_0^{\delta} \sqrt{N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{\mathcal{F}})}\, d\varepsilon$. Let C denote a generic positive and finite constant. It can be different across where it appears. Some empirical processes may not be measurable, and thus the expectation operator cannot be applied to those processes. In such a case, one can replace the expectation operator with the outer expectation operator. We use the notation $\mathbb{E}[\cdot]$ mainly to indicate the expectation operator, but it may also stand for the outer expectation if its argument is not measurable.

a) Proof of Theorem 3.1

Proof. The proof of Theorem 3.1 is analogous to those of Theorem 2 in Lee (2022) and Theorem 4.1 in Lee (2023). To be concrete, we verify the sufficient conditions of Lemma 3.1 in Chen and Pouzo (2015). Assumption 3.1(i) in Chen and Pouzo (2015) is implied by Assumption 2.1. Assumption 3.1(ii) in Chen and Pouzo (2015) is satisfied by Assumptions 3.2 and 3.3. Let

$$Q(\alpha) \equiv \frac{1}{\tau(1-\tau)} \mathbb{E}\left[\left(F_{Y \mid X, Z}\left(g(X, Z_1) + r(X - h(Z)) \mid X, Z\right) - \tau\right)^2\right].$$

Since $F_{Y|X,Z}(\cdot \mid X, Z)$ is continuous almost surely, the function $Q(\alpha)$ is continuous in α, where. Therefore, Assumption 3.1(iii) in Chen and Pouzo (2015) is met. Under Assumption 3.4, $\sum(X, Z) = \tau(1-\tau) > 0$, which implies Assumption 3.1(iv) in Chen and Pouzo (2015).

Note that

$$
\begin{aligned}
&Q(\pi_n \alpha_0) \equiv \\
&\frac{1}{\tau(1-\tau)} \mathbb{E}\left[\left(F_{Y \mid X, Z}\left(\pi_n g_0(X, Z_1) + \pi_n r_0(X - \pi_n h_0(Z)) \mid X, Z\right) - \tau\right)^2\right] \\
&\lesssim \left\|\alpha_0 - \pi_n \alpha_0\right\|_{\mathcal{A}, \infty}
\end{aligned}
$$

Since $\tau = F_{Y \mid X, Z}(g_0(X, Z_1) + r_0(X - h_0(Z)) \mid X, Z)$, $\sup_{(y, x', z')} f_{Y \mid X, Z}(y \mid x, z) < \infty$ and $\left\|\nabla r_0\right\|_\infty < \infty$ under Assumptions 3.1 and 3.2. Together with Assumption 3.4, we have Assumption 3.2(i) in Chen and Pouzo (2015) satisfied. Since $P(\cdot) = \hat{P}_n(\cdot)$, Assumption 3.2 (ii) is satisfied. By the same argument of the proof of Lemma B. 1 in Lee (2023), it follows that Assumption 3.2(iii) in Chen and Pouzo (2015) is satisfied.

We verify the conditions of Lemma C. 2 in Chen and Pouzo (2012) for Assumption 3.3 in Chen and Pouzo (2015). We adapt the proof of Lemma 2 in Lee (2022). Specifically, we can show that

$$
\sup_{\alpha \in \mathcal{A}_n} \left|\rho(W, \alpha)\right| \leq 2
$$

and that

$$
\mathbb{E}\left[\left(b_j(X, Z)\right)^2 \sup_{\tilde{\alpha} \in \mathcal{A}_n : \left\|\alpha - \tilde{\alpha}\right\|_{\mathcal{A}, \infty} \leq \delta} \left\|\rho(W; \alpha) - \rho(W; \tilde{\alpha})\right\|_E^2\right] \leq K^2 \delta
$$

under the assumptions in Theorem 3.1 by using the same argument for the proof of Lemma 2 in Lee (2022). In addition, we have

$$
\begin{aligned}
&\int_0^1 \sqrt{1 + \log N(w^{1/\kappa}, \mathcal{A}, \|\cdot\|_{\mathcal{A}, \infty})}\, dw \\
&\lesssim \int_0^1 \left(w^{-\frac{(d_x + d_{z_1})}{p_g}} + w^{-\frac{d_x}{p_r}} + w^{-\frac{d_z}{p_h}}\right) dw < \infty
\end{aligned}
$$

with $\kappa = 1$ under Assumption 3.2 by the same argument for the proof of Lemma 2 in Lee (2022). Therefore, we have Assumption 3.3 in Chen and Pouzo (2015) in our case with

$$
\bar{\delta}_{m,n}^2 = \eta_{0,n} = \max\left\{\frac{J_n}{n}, J_n^{-\frac{2p_m}{d_x + d_z}}\right\} = o(1).
$$

In all, the conditions of Lemma 3.1 in Chen and Pouzo (2015) are satisfied, and this completes the proof.

b) Proof of Theorem 3.2

Proof. The proof of Theorem 3.2 is almost identical to that of Theorem 3 in Lee (2022), except for that we use a non-trivial penalty function. Therefore, it is enough to verify Assumption 3.4(iii) and (iv) in Chen and Pouzo (2015). They are implied by Assumptions 3.4 and 3.8, and this completes the proof.

c) Proof of Theorem 3.3

Lemma A.1. Suppose that Assumptions in Theorem 3.3 hold. Then, Assumptions A. 4 and A. 5 in Chen and Pouzo (2015) are satisfied.

Proof. The proof for Assumption A. 4 is identical to the proof of Lemma A. 2 in Lee (2022). Assumptions 3.1 and 3.6 imply conditions (i), (ii), and (iv) of Assumption A. 4 in Chen and Pouzo (2015). Condition (iii) of Assumption A. 4 in Chen and Pouzo (2015) can be implied by the condition on J_n in Assumption 3.6 with polynomial spline sieve, as demonstrated in Assumption C. 1 in Chen and Pouzo (2012). Therefore, Assumption A. 4 in Chen and Pouzo (2015) is satisfied.

It is straightforward to see that $\sup_{\alpha \in \mathcal{A}_n} \left|\rho(W, \alpha)\right| \leq 2$ and that

$$\mathbb{E}\left[\sup_{\tilde{\alpha} \in \mathcal{N}_{osn} : \|\alpha - \tilde{\alpha}\|_{\mathcal{A},2} \leq \delta} \left|\rho(W; \alpha) - \rho(W; \tilde{\alpha})\right|^2\right] \leq K\delta \tag{13}$$

for some constant $K > 0$ under Assumptions 3.2 and 3.3. Therefore, conditions (i) and (ii) of Assumption A. 5 in Chen and Pouzo (2015) are satisfied with $\bar{\rho}_n(W) \equiv 2$ and $\kappa = 1/2$.

Let $\sqrt{C_n} \equiv \int_0^1 \sqrt{1 + \log\left(N_{[]}\left(\omega\left(M_n \delta_{2,n}\right)^{1/2}, \mathcal{O}_{on}, \|\cdot\|_2\right)\right)} d\omega$, where

$$\mathcal{O}_{on} \equiv \{\rho(\cdot\,;\,\alpha) - \rho(\cdot\,;\,\alpha_0) : \alpha \in \mathcal{N}_{osn}\}$$

It follows from equation (13) that

$$\log\left(N_{[]}\left(\omega\,\delta_{2,n}^{1/2}, \mathcal{O}_{on}, \|\cdot\|_2\right)\right) \lesssim \log\left(N\left(\omega\,\delta_{2,n}^{1/2}, \mathcal{N}_{osn}, \|\cdot\|_{\mathcal{A},2}\right)\right)$$

by Theorem 9.23 in Kosorok (2008). Therefore, it is enough to bound $\log\left(N\left(\omega\, \delta_{2,n}^{1/2},\, \mathcal{A}_n,\, \|\cdot\|_{\mathcal{A},2}\right)\right)$. Define

$$\mathcal{G}_{osn} \equiv \{g \in \mathcal{G}_n : \|g - g_0\|_2 \leq \delta_{2,n}\},$$
$$\mathcal{R}_{osn} \equiv \{r \in \mathcal{R}_n : \|r - r_0\|_2 \leq \delta_{2,n}\},$$
$$\mathcal{H}_{osn} \equiv \{h \in \mathcal{H}_n : \|h - h_0\|_2 \leq \delta_{2,n}\},$$

then $\mathcal{N}_{osn} \subseteq \mathcal{G}_{osn} \times \mathcal{R}_{osn} \times \mathcal{H}_{osn}$, and we obtain that

$$\log\left(N\left(\omega,\, \mathcal{N}_{osn},\, \|\cdot\|_{\mathcal{A},2}\right)\right) \lesssim \log N_{[]}\left(\omega,\, \mathcal{G}_{osn},\, \|\cdot\|_2\right) + \log N_{[]}\left(\omega,\, \mathcal{R}_{osn},\, \|\cdot\|_2\right) + \log N_{[]}\left(\omega,\, \mathcal{H}_{osn},\, \|\cdot\|_2\right)$$

by Lemma 9.18 in Kosorok (2008). It then follows from Corollary 2.6 in Van de Geer (2000) that

$$\begin{aligned} &\log\left(N\left(\omega\delta_{2,n}^{1/2},\, \mathcal{N}_{osn},\, \|\cdot\|_{\mathcal{A},2}\right)\right) \\ &\lesssim \log N_{[]}\left(\omega\delta_{2,n}^{1/2},\, \mathcal{G}_{osn},\, \|\cdot\|_2\right) + \log N_{[]}\left(\omega\delta_{2,n}^{1/2},\, \mathcal{R}_{osn},\, \|\cdot\|_2\right) + \\ &\quad \log N_{[]}\left(\omega\delta_{2,n}^{1/2},\, \mathcal{H}_{osn},\, \|\cdot\|_2\right) \\ &\lesssim \max\{k_{g,n},\, k_{r,n},\, k_{h,n}\} \cdot \log\left(1 + \frac{C}{\omega\delta_{2,n}^{1/2}}\right) \end{aligned}$$

under Assumption 3.3, and we have

$$\begin{aligned} \sqrt{C_n} &\lesssim \int_0^1 \sqrt{1 + \max\{k_{g,n},\, k_{r,n},\, k_{h,n}\} \cdot \log\left(1 + \frac{C}{\omega\delta_{2,n}^{1/2}}\right)}\, d\omega \\ &\lesssim \sqrt{\max\{k_{g,n},\, k_{r,n},\, k_{h,n}\}}\ \delta_{2,n}^{-1/4} \end{aligned}$$

as $\log(1 + x) \leq x$ for all $x \geq 0$. This leads to that

$$n\left(\delta_{2,n}^{*}\right)^2 \delta_{2,n}^{1/2} \cdot \sqrt{C_n} \cdot M_n \lesssim n\left(\delta_{2,n}^{*}\right)^2 \delta_{2,n}^{1/4} \sqrt{\max\{k_{g,n},\, k_{r,n},\, k_{h,n}\}}\, M_n = o(1)$$

under Assumption 3.11. As a result, condition (iii) of Assumption A. 5 in Chen and Pouzo (2015) is implied. Condition (iv) of Assumption A. 5 in Chen and Pouzo (2015) is implied by Assumptions 3.4 and 3.11. In all, Assumptions A. 4 and A. 5 in Chen and Pouzo (2015) are satisfied.

Let $\tilde{m}(X, Z ; \alpha)$ be the least square projection of $m(X, Z ; \alpha)$ onto $b^{J_n}(X, Z)$. Let $A(X, Z; u_n^*) \equiv \frac{1}{\tau(1-\tau)} \frac{dm(X, Z ; \alpha_0)}{d\alpha}[u_n^*]$ and $\tilde{A}(X, Z ; u_n^*)$ be the least square projection of $A(X, Z ; u_n^*)$ onto $b^{J_n}(X, Z)$.

Lemma A.2. Suppose that Assumptions in Theorem 3.3 hold. Then, Assumption A. 6 in Chen and Pouzo (2015) is satisfied.

Proof. Since $A(X, Z ; u_n^*) \in \Lambda_{\tilde{c}_a}^{p_a}\left(Supp(X, Z)\right)$ with $\tilde{c}_a = \frac{c_a}{\tau(1-\tau)}$ by Assumption 3.10 and $(b_j(\cdot))_{j=1}^{\infty}$ is a sequence of polynomial spline functions by Assumption 3.6, it follows that

$$\left\|\tilde{A}(X, Z ; u_n^*) - A(X, Z ; u_n^*)\right\|_2^2 = O\left(J_n^{-\frac{2p_a}{(d_x + d_z)}}\right)$$

by Newey (1997). Therefore, under Assumptions 3.4 and 3.11,

$$\left\|\tilde{A}(X, Z ; u_n^*) - A(X, Z ; u_n^*)\right\|_2^2 \cdot \delta_{2,n}^2 = o\left(n^{-1}\right),$$

and this implies that conditions (i) and (ii) of Assumption A. 6 in Chen and Pouzo (2015) are satisfied.

Lemma A.3. Suppose that Assumptions in Theorem 3.3. Then, Assumption A. 7 in Chen and Pouzo (2015) is satisfied.

Proof. Since $F_{Y\mid X, Z}(y|x, z)$ is twice continuously differentiable with respect to y for almost all $(x', z') \in Supp(X, Z)$ by Assumption 3.10, condition (i) of Assumption A. 7 in Chen and Pouzo (2015) is met. For any $\alpha \in \mathcal{N}_{osn}$, it follows that

$$\left|\frac{dm(X, Z ; \alpha)}{d\alpha}[u_n^*] - \frac{dm(X, Z ; \alpha_0)}{d\alpha}[u_n^*]\right| = \left|\frac{d^2m(X, Z ; \tilde{\alpha})}{d\alpha^2}[u_n^*, \alpha - \alpha_0]\right|,$$

where $\tilde{\alpha}$ lies between α and α_0, by the mean-value theorem. Note that

$$\frac{d^2m(X, Z ; \tilde{\alpha})}{d\alpha^2}[u_n^*, v]$$
$$= f'_{Y\mid X, Z}\left(\tilde{g}(X) + \tilde{r}(X - \tilde{h}(Z)) \mid X, Z\right) \cdot \{u_{g,n}^* + u_{r,n}^* + \nabla r_0 \cdot u_{h,n}^*\} \cdot$$
$$\{v_g + v_r + \nabla r_0 \cdot v_h\}.$$

Since $f'_{Y|X, Z}(\cdot|X, Z)$ is uniformly bounded by Assumption 3.10, we obtain that

$$\mathbb{E}\left[\sup_{\alpha \in \mathcal{N}_{osn}} \left|\frac{dm(X, Z; \alpha)}{d\alpha}[u_n^*] - \frac{dm(X, Z; \alpha_0)}{d\alpha}[u_n^*]\right|\right]$$
$$\lesssim \delta_{2,n}\sqrt{\mathbb{E}\left[\{u_{g,n}^* + u_{r,n}^* + \nabla r_0 \cdot u_{h,n}^*\}^2\right]} = O(\delta_{2,n}).$$

Therefore, Assumption 3.11 implies that

$$\mathbb{E}\left[\sup_{\alpha \in \mathcal{N}_{osn}} \left|\frac{dm(X, Z; \alpha)}{d\alpha}[u_n^*] - \frac{dm(X, Z; \alpha_0)}{d\alpha}[u_n^*]\right|\right] \cdot \delta_{2,n}^2 = o(n^{-1}),$$

which implies condition (ii) of Assumption A. 7 in Chen and Pouzo (2015). In addition, Assumption 3.10 implies that

$$\mathbb{E}\left[\sup_{\alpha \in \mathcal{N}_{osn}} \left|\frac{d^2 m(X, Z; \alpha)}{d\alpha^2}[u_n^*, u_n^*]\right|\right] \cdot \delta_{2,n}^2 = o(1)$$

since $\mathbb{E}\left[\sup_{\alpha \in \mathcal{N}_{osn}} \left|\frac{d^2 m(X, Z; \alpha)}{d\alpha^2}[u_n^*, u_n^*]\right|\right] = O(1)$, and this leads to that condition (iii) of Assumption A. 7 in Chen and Pouzo (2015) is satisfied. Lastly, for any $\alpha_1 \in \mathcal{N}_{os}$ and $\alpha_2 \in \mathcal{N}_{osn}$,

$$\left|\frac{dm(X, Z; \alpha_1)}{d\alpha}[\alpha_2 - \alpha_0] - \frac{dm(X, Z; \alpha_0)}{d\alpha}[\alpha_2 - \alpha_0]\right| \leq$$
$$\sup_{\alpha \in \mathcal{N}_{os}} \left|\frac{d^2 m(X, Z; \alpha)}{d\alpha^2}[\alpha_2 - \alpha_0, \alpha_1 - \alpha_0]\right|$$
$$\lesssim \sup_{\alpha_1 \in \mathcal{N}_{os}} \|\alpha_1 - \alpha_0\|_E \cdot \sup_{\alpha_2 \in \mathcal{N}_{osn}} \|\alpha_2 - \alpha_0\|_E,$$

where the second line holds by Assumption 3.10 and the Cauchy-Schwarz inequality. Therefore, it follows that

$$\left|\mathbb{E}\left[\left(A(X, Z; u_n^*) \cdot \frac{dm(X, Z; \alpha_1)}{d\alpha}[\alpha_2 - \alpha_0]\right.\right.\right.$$
$$\left.\left.\left. - \frac{dm(X, Z; \alpha_0)}{d\alpha}[\alpha_2 - \alpha_0]\right)\right|\right. \leq \sqrt{\mathbb{E}\left[\left(A(X, Z; u_n^*)\right)^2\right]}$$

$$\cdot \sqrt{\mathbb{E}\left[\left(\frac{dm(X, Z ; \alpha_1)}{d\alpha}[\alpha_2 - \alpha_0] - \frac{dm(X, Z ; \alpha_0)}{d\alpha}[\alpha_2 - \alpha_0]\right)^2\right]}$$
$$\sqrt{\mathbb{E}\left[\left(\sup_{\alpha \in \mathcal{N}_{os}}\left|\frac{d^2 m(X, Z ; \alpha)}{d\alpha^2}[\alpha_2 - \alpha_0, \alpha_1 - \alpha_0]\right|\right)^2\right]}$$
$$\sqrt{\mathbb{E}\left[\sup_{\alpha_1 \in \mathcal{N}_{os}}\|\alpha_1 - \alpha_0\|_E^2 \cdot \sup_{\alpha_2 \in \mathcal{N}_{osn}}\|\alpha_2 - \alpha_0\|_E^2\right]}$$
$$\leq \delta_{2,n}^2,$$

and $\delta_{2,n}^2 = o(n^{-1/2})$ by Assumption 3.11, implying that condition (iv) of Assumption A. 7 in Chen and Pouzo (2015) is satisfied. This completes the proof.

Proof of the theorem

Proof. Assumptions 3.1-3.4 in Chen and Pouzo (2015) are implied by Assumptions 2.1-3.7, as shown in Lee (2022). Assumption 3.5 in Chen and Pouzo (2015) is directly imposed by Assumption 3.9. Lemmas A.1, A.2, and A. 3 imply that under the set of assumptions of Theorem 3.3, Assumptions A4-A7 in Chen and Pouzo (2015) are met. Applying part (1) of Lemma 5.1 in Chen and Pouzo (2015) yields that Assumption 3.6(i) in Chen and Pouzo (2015) is satisfied. Under Assumption 3.12, it can be easily shown that for any $\varepsilon > 0$,

$$\mathbb{E}\left[\left(\frac{S_{n,i}^*}{\|v_n^*\|_{sd}}\right)^2 \mathbf{1}\left(\left|\frac{S_{n,i}^*}{\|v_n^*\|_{sd}}\right| > \sqrt{n}\varepsilon\right)\right]$$
$$= \mathbb{E}\left[\left(\frac{S_{n,i}^*}{\|v_n^*\|_{sd}}\right)^{2+\delta_0}\left(\frac{S_{n,i}^*}{\|v_n^*\|_{sd}}\right)^{-\delta_0} \mathbf{1}\left(\left|\frac{S_{n,i}^*}{\|v_n^*\|_{sd}}\right| > \sqrt{n}\varepsilon\right)\right]$$
$$\leq \mathbb{E}\left[\left(\frac{S_{n,i}^*}{\|v_n^*\|_{sd}}\right)^{2+\delta_0} \frac{1}{\left(\sqrt{n}\varepsilon\right)^{\delta_0}}\right]$$
$$= o(1)$$

which is the Lindeberg condition for Lindeberg's central limit theorem. As a result, Assumption 3.6(ii) in Chen and Pouzo (2015) is satisfied. Therefore, all conditions in Theorem 4.1 in Chen and Pouzo (2015) are satisfied, and this completes the proof.

d) Proof of Theorem 3.4

Proof. Note that under Assumption 3.4, the PSMD estimator defined in (6) is the optimally weighted one. Therefore, the result is implied by Theorem 4.3 in Chen and Pouzo (2015) and the proof of Theorem 3.3.